

# Sentiment Analysis based on Psychological and Linguistic Features for Spanish language

María del Pilar Salas-Zárate<sup>1</sup>, Mario Andrés Paredes-Valverde<sup>1</sup>, Miguel Ángel Rodríguez-García<sup>2</sup>, Rafael Valencia-García<sup>1</sup>, Giner Alor-Hernández<sup>3</sup>

<sup>1</sup>Departamento de Informática y Sistemas, Universidad de Murcia. Campus de Espinardo 30100 Murcia, Spain. E-mail:{mariapilar.salas, valencia}@um.es

<sup>2</sup>Computational Bioscience Research Center, King Abdullah University of Science and Technology, 4700 KAUST, P.O. Box 2882, 23955-6900 Thuwal, Kingdom of Saudi Arabia.

<sup>3</sup>Division of Research and PostgraduateStudies Instituto Tecnológico de Orizaba, Av. Oriente 9, No. 852, Col. E. Zapata, CP 94320 Orizaba, Veracruz, Mexico. E-mail: galor@itorizaba.edu.mx

**Abstract** Recent research activities in the areas of opinion mining, sentiment analysis and emotion detection from natural language texts are gaining ground under the umbrella of affective computing. Nowadays, there is a huge amount of text data available in the Social Media (e.g. forums, blogs, and social networks) concerning to users' opinions about experiences buying products and hiring services. Sentiment analysis or opinion mining is the field of study that analyses people's opinions and mood from written text available on the Web. In this paper, we present extensive experiments to evaluate the effectiveness of the psychological and linguistic features for sentiment classification. To this purpose, we have used four psycholinguistic dimensions obtained from LIWC, and one stylometric dimension obtained from WordSmith, for the subsequent training of the SVM, Naïve Bayes, and J48 algorithms. Also, we create a corpus of tourist reviews from the travel website TripAdvisor. The findings reveal that the stylometric dimension is quite feasible for sentiment classification. Finally, with regard to the classifiers, SVM provides better results than Naïve Bayes and J48 with an F-measure rate of 90.8%.

## 1 Introduction

With the booming of social media, some applications using sentiment analysis has been rapidly developed in recent years. The majority of people visit opinion sites, discussion forums or social networks aiming to obtain experiences of other users before taking a decision. Due to the number of reviews has exponentially

increased on the Web, the task of reading all these opinions has become impossible for the users. For this reason, there is a need for automating the analysis and classification of these opinions.

Sentiment analysis (a.k.a. opinion mining) has become a popular topic in order to understand public opinion from unstructured Web data. From this perspective, sentiment analysis is devoted to extract users' opinions from textual data. The capture of public opinion is gaining momentum, particularly in terms of product preferences, marketing campaigns, political movements, financial aspects and company strategies. In this context, different techniques based on supervised learning and unsupervised learning, may be used. Thanks to these techniques, several attempts at sentiment classification were done. For instance, the problems that have been studied are mainly focused around subjectivity identification [1], analysis and building sentiment lexicons [2] [3], evaluation and classification of the Twitter messages [4] [5] [6], features selection [7], negation [8]. Some other proposals have tried to introduce sentiment classification problem in different levels i.e. document-level [9] [10], sentence-level [11], word-level and aspect-level [12]. However, one of the main issues on sentiment classification is the existence of many conceptual rules that govern the linguistic expression of sentiments. Human psychology, which relates to social, cultural, personal and biological aspects, can be an important feature to be considered in sentiment analysis. For this purpose, the LIWC and WordSmith text analysis software are useful tools since they enable the extraction of psychological and linguistic features from natural language text.

On the other hand, it is worth mentioning that most sentiment analysis studies deal exclusively with English documents, perhaps owing to the lack of resources in other languages. Considering that the Spanish language has a much more complex syntax than many other languages, and that it is the third most widely spoken language in the world, we firmly believe that the computerization of Internet domains in this language is of utmost importance.

The main contribution of this work is the evaluation of the effectiveness of different psychological and linguistic features for sentiment classification in the Spanish language. This process covered five categories (positive, negative, neutral, highly positive and highly negative) and it was performed by using different classifiers. Also, a corpus of tourism reviews, obtained from the travel website TripAdvisor, was built and compiled.

This paper is structured as follows. Section 2 presents the state of the art on opinion mining and sentiment analysis. Section 3 presents the tourism corpus used along the evaluation performed. Section 4 describes the LIWC's dimensions and stylometric variables provided by Wordsmith. Section 5 presents the classification algorithms SVM (Support Vector Machine), Naïve Bayes, and J48 used in this work. Section 6 provides a detailed description about the experiments performed. Section 7 presents the results of the classifiers' evaluation performed by using LIWC's dimensions and stylometric variables described in section 4 and 5. Section 8 discusses the results obtained by our approach, and a comparison is

presented with two machine-learning based natural language processing (NLP) tools, the OpenNLP and Stanford Classifier. Finally, Section 9 describes conclusions and future work.

## 2 Related Work

In recent years, several proposals have presented studies based on two main approaches in the polarity classification: Semantic Orientation and Machine learning.

The Semantic Orientation (SO) approach uses lexical resources such as lexicons, which have been automatically or semi-automatically generated [13], [14], [15]. These lexicons have usually extended the WordNet lexical database. Two clear examples of this kind of lexicons are WordNet-Affect[14] and SentiWordNet [13].

In the literature, several proposals based on sentiment lexicons have been presented. For instance, in [16] a lexicon-based approach to extract sentiment from texts was proposed. This approach is based on The Semantic Orientation CALculator (SO-CAL), which uses dictionaries of words annotated with their semantic orientation (polarity and strength), and incorporates intensification and negation.

[12] proposed an innovative opinion mining methodology that takes advantage of new semantic Web-guided solutions to enhance the results obtained with traditional natural language processing techniques and sentiment analysis processes and Semantic Web technologies. Concretely, the proposed approach is based on three different stages: 1) an ontology based mechanism for feature identification, 2) a technique to assign a polarity to each feature based on SentiWordNet and 3) a new approach for opinion mining based on vector analysis. [17] presented a rule-based method to identify the sentiment polarity of opinion sentences. SentiWordNet was used to calculate overall sentiment score of each sentence. Also, the results showed that SentiWordNet could be used as an important resource for sentiment classification tasks.

Most studies on opinion mining have been concerned to the English language. Indeed, a main disadvantage of the approach based on sentiment lexicons is the lack of sentiment lexicons for non-English languages, therefore it is difficult its application to other languages. Despite the aforementioned issue, there are some interesting works that provide sentiment lexicons oriented to languages such as Spanish [15], [18], German [19], Dutch [20] and Arabic [21].

Also, due to the limitation mentioned above, some proposals have used alternative methods. The first alternative method uses some lexical resources for sentiment analysis based on WordNet (regardless of the target language, for example, the English language) jointly with a lexical database based on WordNet and specific for the target language or multilingual [22].

The second alternative method uses a different sentiment lexicon from the target language. Therefore, an automatic translation of the corpus is carried out before of the polarity classification phase. However, this method depends on the availability and reliability of the automatic translation engine available [23] [24].

On the other hand, there are some proposals [25], [26], [27], [28] based on psycholinguistic tools for sentiment analysis such as LIWC, a tool that offers a dictionary in several languages such as: Spanish, English, French, and German, among others. These proposals construct a semantic orientation-based lexicon base on two ("Positive emotion" and "Negative emotion") of the 76 categories of LIWC. These categories include words such as love, nice, good, great, hurt, ugly, sad, bad, and worse, among others. The LIWC's dictionaries are used on different studies covering languages such as English [25], [27], [28], Portuguese [26] and Spanish [29].

Furthermore, with regard to the Machine learning approach, it is worth mentioning that this approach often relies on supervised classification approaches. These approaches use a collection of data to train the classifier algorithms. Among the machine learning techniques commonly used in the sentiment polarity classification are Support Vector Machine (SVN), Naive Bayes (NB), Maximum Entropy (MaxEnt), among others. For example, in [30] compared three supervised machine learning algorithms (Naive Bayes, SVM and the character based N-gram model) for sentiment classification. This comparison was performed on user's reviews obtained from travel blogs. Authors concluded that well-trained machine learning algorithms can provide a very good performance for sentiment polarity classification on the travel domain. [31] explored how different settings (n-gram size, corpus size, number of sentiment classes, balanced vs. unbalanced corpus, various domains) affect the precision of the machine learning algorithms. Naive Bayes, Decision Tree, and Support Vector Machines were considered.

It is necessary to remark that this work differs from the works presented above in several aspects: (i) our study consider all LIWC's categories for the sentiment analysis. Although the LIWC dictionary has been evaluated in different works for the sentiment analysis, only the "Positive emotion" and "negative emotion" categories have been considered for constructing a lexicon based on semantic orientation, (ii) we have evaluated the combination of psychological and linguistic features extracted by using the LIWC and WordSmith text analyzers, (iii) we carried out a comparison of our approach with two sentiment analysis tools (Stanford Classifier and OpenNLP) based on machine learning, and (iv) unlike of several works, which use only two classes (positive and negative), our approach uses five classes (positive, negative, neutral, highly positive and highly negative). This was done in order to find out how the number of classes affects the precision of the classification algorithms.

### 3 Corpus

For this study, we have chosen the TripAdvisor® website since is the world's largest travel site with 315 million unique users per month and over 200 million reviews and opinions of over 4.4 million accommodations, restaurants and attractions.

The corpus consists of 1600 reviews, obtained from the TripAdvisor® website, concerning hotels, restaurants, and museums, among other topics. Each review was examined and classified by hand in order to ensure the quality of the corpus. Also, a value, from 1 to 5, was assigned to each review according to next criteria:

- 5, when the review has a highly positive sentiment.
- 4, when the review has a positive sentiment.
- 3, when the review has no sentiment.
- 2, when the review has a negative sentiment.
- 1, when the review has a highly negative sentiment.

The classification results were stored in an XML-based format. FIG. 1 shows an example of a review with highly positive orientation.

```
<review id="3" rank="5" topic="Alojamiento">
  <abstract>
    Gran hallazgo
  </abstract>
  <content>
    Fantástico hotel, personal muy amable, habitaciones espaciosas, desayuno excelente
    y muy silencioso para descansar. El vecindario es uno de los mejores de Barna, el
    jardín en la zona posterior es una delicia. Muy pero muy recomendable.
  </content>
</review>
```

**Fig. 1.** Example of a review with highly positive orientation.

Table 1 shows the total amount of reviews classified in each category mentioned above. It should be mentioned that the aforementioned activities were performed along six months by a group of three experts in order to ensure the quality of the corpus.

Value	Classification	Number of reviews
5	Highly positive	320
4	Positive	320
3	Neutral	320
2	Negative	320
1	Highly negative	320
	Total	1600

**Table 1.**Classification of the corpus.

In order to find out how the number of classes affects the precision of the classification algorithms, the corpus was also classified into three categories (Positive, Negative and Neutral). Based on this understanding, the reviews with a 4 and 5 value were classified as positive. Meanwhile, the reviews with a 1 and 2 value were classified as negative. The results of this classification are shown in Table 2.

Value	Classification	Number of reviews
5	Positive	640
4		
3	Neutral	320
2	Negative	640
1		
	Total	1600

**Table 2.**Reviews classified into three categories.

Finally, in order to evaluate the classifier algorithms with only two categories, the neutral opinions obtained from the aforementioned process were omitted. The results of this process are shown in Table 3.

Value	Classification	Number of reviews
5	Positive	640
4		
2	Negative	640
1		
	Total	1280

**Table 3.**Reviews classified into two categories.

#### 4 LIWC and stylometric variables

LIWC (Linguistic Inquiry and Word Count) is a software application that provides an effective tool for studying the emotional, cognitive, and structural components contained in language on a word-by-word basis. Early approaches to psycholinguistic concerns involved almost exclusively qualitative philosophical analyses. In this field, more modern researches have provided empirical evidence

on the relation between language and the state of mind of subjects, or even their mental health. In this regard, further studies such as [32] have dealt with the therapeutic effect of verbally expressing emotional experiences and memories. LIWC was developed for providing an efficient method for studying these psycholinguistic concerns thanks to corpus analysis, and it has been considerably improved since its first version [33]. An updated revision of the original application was presented in [34], namely LIWC2001.

For this study, we used the LIWC2007 Spanish dictionary, which is composed by 7,515 words and word stems. Each word is classified into one or more of the 72 categories included by default in LIWC. Also, these categories are classified into four dimensions: (1) standard linguistic process, (2) psychological process, (3) relativity, and (4) personal concerns.

Table 4 shows some examples of the LIWC categories. The full list of categories is presented in [35].

Dimension	Categories
1.Linguistic processes	Word count, negations, articles, prepositions, among others.
2.Psychological processes	Positive emotions, negative emotions, feeling, communication, among others.
3.Relativity	Past tense verb, future tense verb, inclusive, among others.
4.Personal concerns	Job or work, sports, music, religion, among others.

**Table 4.**LIWC dimensions.

On the other hand, we have used further stylometric variables, due to that there are some linguistic features not included in LIWC standard linguistic dimensions, which has been considered significant for this study. We used word level measures such as: word length distribution, average number words per sentence, word length distribution containing frequency of 1 letter word to frequency of 14 letters word. These linguistic features were obtained from the WordSmith software.

## 5 Machine learning approach

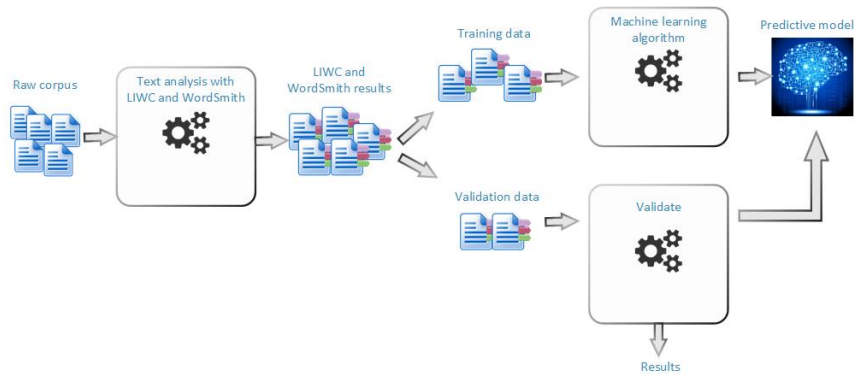
In a classification-based machine learning approach, two sets of data are required: training and validation set. The first set is used by an automatic classifier to learn the differentiating characteristics of documents, and the second set is used to validate the performance of the automatic classifier. Machine learning classifiers such as Naive Bayes (NB), Maximum Entropy (ME), and Support Vector Machines (SVM) have been used on different works achieving great success in text categorization [36], [37], [38].

For this work, we used WEKA [39] to evaluate the classification success of reviews (positive, negative, neutral, highly positive or highly negative) based on LIWC and WordSmith categories.

WEKA provides several classification algorithms (also called classifiers), which allows the creation of models according to the data and purpose of analysis. Classification algorithms are categorized into seven groups: Bayesian (Naïve Bayes, Bayesian nets, etc.), functions (linear regression, SMO, logistic, etc.), lazy (IBk, LWL, etc.), meta-classifiers (Bagging, Vote, etc.), miscellaneous (SerializedClassifier and InputMappedClassifier), rules (DecisionTable, OneR, etc.) and trees (J48, RandomTree, etc.). The classification process involves the building of a model based on the analysis of the instances. This model is represented through classification rules, decision trees, or mathematical formulae. The model is used to generate the classification of unknown data, calculating the percentage of instances which were correctly classified.

## 6 Experiment

This section presents the experiment conducted to study the effectiveness of the psychological and linguistic features for sentiment classification into five categories: highly positive, highly negative, positive, negative and neutral. As it will be shown in next sections, the approach has been tested on the Spanish language by using a corpus of reviews concerning to the tourism domain. Figure 2 shows the general process of the experiment.



**Fig. 2.**General process of the experiment.



### ***6.1 Combination of LIWC dimensions and stylometric dimension***

Aiming to detect the set of dimensions that provide the best results and to validate if the stylometric dimension improves the success rates, we have obtained 11 combinations of LIWC dimensions, and 15 for all possible combinations of stylometric dimension and LIWC dimensions (see Table 5).

Dimensions	Description
1	Dimension 1 LIWC
2	Dimension 2 LIWC
3	Dimension 3 LIWC
4	Dimension 4 LIWC
1-2	Combination of LIWC dimensions 1 and 2
1-3	Combination of LIWC dimensions 1 and 3
1-4	Combination of LIWC dimensions 1 and 4
2-3	Combination of LIWC dimensions 2 and 3
2-4	Combination of LIWC dimensions 2 and 4
3-4	Combination of LIWC dimensions 1 and 2
1-2-3	Combination of LIWC dimensions 1, 2 and 3
1-2-4	Combination of LIWC dimensions 1, 2 and 4
1-3-4	Combination of LIWC dimensions 1, 3 and 4
2-3-4	Combination of LIWC dimensions 2, 3 and 4
1-2-3-4	Combination of all LIWC dimensions
Styl.	Further stylometric variables
1+ Styl.	Combination LIWC dimension 1 and Further stylometric variables
2+ Styl.	Combination LIWC dimension 2 and Further stylometric variables
3+ Styl.	Combination LIWC dimension 3 and Further stylometric variables
4+ Styl.	Combination LIWC dimension 4 and Further stylometric variables
1_2+ Styl.	Combination LIWC dimension 1, 2 and Further stylometric variables
1_3+ Styl.	Combination LIWC dimension 1, 3 and Further stylometric variables
1_4+ Styl.	Combination LIWC dimension 1, 4 and Further stylometric variables
2_3+ Styl.	Combination LIWC dimension 2, 3 and Further stylometric variables
2_4+ Styl.	Combination LIWC dimension 2, 4 and Further stylometric variables
3_4+ Styl.	Combination LIWC dimension 3, 4 and Further stylometric variables
1_2_3+ Styl.	Combination LIWC dimension 1, 2, 3 and Further stylometric variables
1_2_4+ Styl.	Combination LIWC dimension 1, 2, 4 and Further stylometric variables
1_3_4+ Styl.	Combination LIWC dimension 1, 3, 4 and Further stylometric variables
2_3_4+ Styl.	Combination LIWC dimension 2, 3, 4 and Further stylometric variables
1_2_3_4+ Styl.	Combination all LIWC dimension and Further stylometric variables

**Table 5.**Combination of LIWC dimensions and stylistometric dimension.

## ***6.2 Text analysis with LIWC and WordSmith***

The tourism corpus was analyzed with LIWC and WordSmith software through all the possible combinations presented in Table 5 and taking into account three possible sets of opinion classes (positive-negative, positive-neutral-negative and highly positive-positive-neutral-negative-highly negative).

LIWC searches for target words or word stems from the dictionary, categorizes them into one of its linguistic dimensions, and then converts the raw counts to percentages of total words. On the other hand, WordSmith generates statistical information such as average sentence and average word length, among others.

## ***6.3 Training a machine learning algorithm and validation test***

A set of the results obtained by the analysis with LIWC and WordSmith was used in order to train the machine learning algorithms, which are subsequently validated with the remaining subset. For the experiment, J48, Naive Bayes (NB), and the SMO algorithm for SVM classifiers were used.

Specifically, for J48, NB, and SMO classifiers, a ten-fold cross-validation was done. This technique is used to evaluate how the results obtained would generalize to an independent data set. Since the aim of this experiment is the prediction of the positive, negative, neutral, highly positive and highly negative condition of the texts, a cross-validation was applied in order to estimate the accuracy of the predictive models. It involved partitioning a sample of data into complementary subsets, performing an analysis on the training set and validating the analysis on the testing or validation set.

## **7 Evaluation and Results**

In order to measure the performance of our method, we have used three evaluation measurements that are commonly used in sentiment analysis: precision, recall and F-measure. Recall (1) is the proportion of actual positive cases that were correctly predicted as such. On the other hand, precision (2) represents the proportion of predicted positive cases that are real positives. Finally, F-measure (3) is the harmonic mean of precision and recall.

$$(1) \text{ Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$(2) \text{ Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$(3) \text{ F1} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

### 7.1 Results for the Tourism Corpus

This section describes the results obtained of each possible combination described in the previous section. Table 6, 7 and 8 present the evaluation results by using two (positive-negative), three (positive-neutral-negative) and five (highly positive-positive-neutral-negative-highly negative) categories respectively. In the first column, the number of LIWC dimensions used for each classifier is indicated. For instance, 1\_2\_3\_4 indicates that all the dimensions have been used in the experiment, and 1\_2 indicates that only the 1 and 2 dimensions have been used to train the classifier.

	J48			Naïve Bayes			SMO		
	P	R	F1	P	R	F1	P	R	F1
1	0.83	0.83	0.83	0.809	0.802	0.799	0.825	0.823	0.823
2	0.788	0.788	0.787	0.808	0.791	0.784	0.836	0.834	0.834
3	0.737	0.735	0.734	0.728	0.702	0.68	0.736	0.736	0.736
4	0.667	0.661	0.651	0.655	0.654	0.652	0.673	0.66	0.633
1_2	0.806	0.805	0.805	0.838	0.833	0.831	0.898	0.898	0.898
1_3	0.799	0.798	0.798	0.804	0.804	0.804	0.833	0.832	0.832
1_4	0.8	0.799	0.799	0.785	0.78	0.778	0.836	0.834	0.833
2_3	0.803	0.803	0.803	0.8	0.787	0.781	0.852	0.851	0.85
2_4	0.8	0.798	0.798	0.787	0.782	0.78	0.837	0.836	0.836
3_4	0.744	0.744	0.743	0.693	0.69	0.687	0.735	0.734	0.734
1_2_3	0.821	0.821	0.821	0.829	0.822	0.82	0.905	0.905	0.905
1_2_4	0.819	0.819	0.819	0.831	0.829	0.828	0.905	0.905	0.905
1_3_4	0.785	0.784	0.784	0.772	0.772	0.772	0.829	0.828	0.828
2_3_4	0.767	0.767	0.767	0.79	0.784	0.782	0.851	0.85	0.85
1_2_3_4	0.818	0.818	0.818	0.825	0.822	0.821	0.903	0.903	0.903
Styl.	0.697	0.697	0.697	0.759	0.719	0.693	0.749	0.743	0.739
1+ Styl.	0.77	0.77	0.77	0.788	0.756	0.741	0.828	0.824	0.823

	J48			Naïve Bayes			SMO		
2+ Styl.	0.772	0.772	0.772	0.788	0.775	0.799	0.86	0.859	0.859
3+ Styl.	0.718	0.717	0.717	0.758	0.725	0.704	0.772	0.772	0.772
4+ Styl.	0.688	0.688	0.688	0.753	0.72	0.697	0.752	0.751	0.75
1_2+ Styl.	0.81	0.81	0.81	0.829	0.819	0.816	0.902	0.902	0.902
1_3+ Styl.	0.766	0.766	0.765	0.796	0.768	0.756	0.83	0.828	0.828
1_4+ Styl.	0.772	0.772	0.772	0.783	0.756	0.743	0.833	0.83	0.829
2_3+ Styl.	0.773	0.773	0.773	0.794	0.785	0.782	0.869	0.869	0.869
2_4+ Styl.	0.785	0.785	0.785	0.785	0.773	0.767	0.859	0.859	0.859
3_4+ Styl.	0.706	0.706	0.706	0.753	0.726	0.708	0.774	0.773	0.773
1_2_3+ Styl.	0.816	0.816	0.816	0.83	0.822	0.819	0.906	0.906	0.906
1_2_4+ Styl.	0.798	0.798	0.798	0.82	0.809	0.806	0.907	0.907	0.907
1_3_4+ Styl.	0.765	0.765	0.765	0.787	0.764	0.754	0.829	0.828	0.828
2_3_4+ Styl.	0.788	0.788	0.788	0.791	0.783	0.779	0.865	0.864	0.864
1_2_3_4+ Styl.	0.806	0.806	0.806	0.822	0.813	0.811	0.908	0.908	0.908

**Table 6.** Classification results obtained by using two classes (positive-negative).

Table 6 shows that the SMO algorithm provides better results than J48 and Naïve Bayes. When the evaluation involved only one dimension, the J48 and Naïve algorithms obtained the best results through the first dimension (Linguistic processes). On the other hand, the SMO algorithm obtained the best result through the second dimension (Psychological processes). Conversely, the fourth dimension (Personal concerns) provided the worst results for the three classification algorithms. On the other hand, the combination of all LIWC dimensions and the stylometric dimension provided the best classification result with an F-measure of 90.8%. As can be seen, the addition of the stylometric dimension improved the classification results.

	J48			Naïve Bayes			SMO		
	P	R	F1	P	R	F1	P	R	F1
1	0.724	0.744	0.731	0.712	0.741	0.713	0.66	0.774	0.71
2	0.679	0.693	0.685	0.708	0.732	0.7	0.675	0.793	0.727
3	0.655	0.689	0.653	0.631	0.66	0.611	0.604	0.704	0.648
4	0.619	0.638	0.598	0.615	0.614	0.614	0.581	0.653	0.583
1_2	0.715	0.721	0.718	0.746	0.763	0.748	0.792	0.842	0.776
1_3	0.698	0.702	0.7	0.713	0.724	0.718	0.667	0.783	0.718
1_4	0.706	0.718	0.708	0.714	0.699	0.702	0.668	0.784	0.718
2_3	0.697	0.702	0.699	0.704	0.714	0.691	0.684	0.804	0.737

	J48			Naïve Bayes			SMO		
2_4	0.684	0.689	0.686	0.716	0.711	0.707	0.672	0.789	0.724
3_4	0.636	0.646	0.64	0.631	0.616	0.619	0.614	0.716	0.658
1_2_3	0.724	0.721	0.722	0.742	0.746	0.739	0.798	0.842	0.78
1_2_4	0.716	0.72	0.718	0.756	0.748	0.749	0.802	0.845	0.784
1_3_4	0.696	0.701	0.698	0.714	0.691	0.7	0.672	0.789	0.723
2_3_4	0.682	0.688	0.684	0.708	0.694	0.694	0.706	0.804	0.74
1_2_3_4	0.725	0.73	0.73	0.758	0.744	0.747	0.775	0.839	0.783
Styl.	0.625	0.632	0.628	0.663	0.681	0.625	0.616	0.714	0.654
1+ Styl.	0.674	0.678	0.676	0.687	0.691	0.657	0.662	0.776	0.711
2+ Styl.	0.679	0.682	0.68	0.696	0.713	0.69	0.688	0.809	0.742
3+ Styl.	0.621	0.628	0.624	0.666	0.678	0.633	0.632	0.741	0.681
4+ Styl.	0.616	0.623	0.619	0.678	0.673	0.644	0.621	0.724	0.665
1_2+ Styl.	0.721	0.716	0.718	0.731	0.738	0.724	0.822	0.843	0.779
1_3+ Styl.	0.653	0.656	0.654	0.703	0.698	0.675	0.669	0.786	0.721
1_4+ Styl.	0.678	0.677	0.677	0.693	0.681	0.662	0.673	0.789	0.724
2_3+ Styl.	0.678	0.674	0.676	0.706	0.719	0.704	0.691	0.813	0.745
2_4+ Styl.	0.68	0.679	0.679	0.71	0.703	0.696	0.686	0.807	0.74
3_4+ Styl.	0.622	0.622	0.622	0.679	0.671	0.653	0.631	0.739	0.679
1_2_3+ Styl.	0.726	0.723	0.725	0.733	0.734	0.726	0.791	0.837	0.779
1_2_4+ Styl.	0.717	0.718	0.717	0.732	0.721	0.718	0.789	0.841	0.783
1_3_4+ Styl.	0.666	0.664	0.665	0.704	0.686	0.675	0.674	0.79	0.725
2_3_4+ Styl.	0.663	0.662	0.663	0.714	0.706	0.704	0.731	0.812	0.749
1_2_3_4+ Styl.	0.724	0.726	0.725	0.743	0.729	0.73	0.803	0.844	0.792

**Table 7.** Classification results obtained by using three classes (positive-neutral-negative).

The classification results obtained by using three classes are presented in Table 7. We found that the SMO classification algorithm provides the best results. Individually, the first dimension (Linguistic processes) provided the best results for the J48 and Naïve Bayes algorithms. On the other hand, the second dimension (Psychological processes) provided the best results for the SMO algorithm. Quite the reverse, the third and four dimensions provide the worst results. Furthermore, the combination of all LIWC dimensions and the stylometric dimension provided the best classification result with an F-measure of 79.2%. As can be seen, in these results also the addition of the stylometric dimension improved the classification results.

	J48	Naïve Bayes	SMO
--	-----	-------------	-----

	J48			Naïve Bayes			SMO		
	P	R	F1	P	R	F1	P	R	F1
1	0.558	0.558	0.558	0.579	0.586	0.572	0.611	0.61	0.603
2	0.566	0.566	0.566	0.58	0.588	0.562	0.607	0.615	0.605
3	0.534	0.534	0.534	0.529	0.539	0.507	0.537	0.557	0.522
4	0.54	0.539	0.538	0.52	0.518	0.498	0.538	0.533	0.512
1_2	0.598	0.599	0.598	0.59	0.603	0.582	0.653	0.66	0.65
1_3	0.578	0.579	0.578	0.572	0.588	0.568	0.609	0.611	0.607
1_4	0.561	0.561	0.561	0.566	0.559	0.549	0.61	0.606	0.603
2_3	0.589	0.591	0.59	0.58	0.586	0.564	0.624	0.633	0.622
2_4	0.585	0.586	0.586	0.577	0.574	0.568	0.61	0.616	0.609
3_4	0.536	0.537	0.536	0.529	0.522	0.515	0.543	0.554	0.542
1_2_3	0.577	0.575	0.575	0.586	0.599	0.579	0.636	0.644	0.637
1_2_4	0.575	0.576	0.575	0.605	0.603	0.601	0.647	0.652	0.646
1_3_4	0.566	0.566	0.566	0.566	0.563	0.559	0.6	0.601	0.598
2_3_4	0.575	0.576	0.575	0.583	0.578	0.574	0.621	0.629	0.62
1_2_3_4	0.601	0.603	0.602	0.594	0.596	0.592	0.645	0.651	0.645
Styl.	0.539	0.539	0.539	0.509	0.543	0.479	0.551	0.564	0.549
1+ Styl.	0.59	0.59	0.59	0.543	0.561	0.521	0.61	0.609	0.601
2+ Styl.	0.575	0.576	0.576	0.556	0.584	0.55	0.606	0.619	0.608
3+ Styl.	0.543	0.544	0.544	0.501	0.541	0.496	0.556	0.575	0.555
4+ Styl.	0.543	0.543	0.543	0.553	0.543	0.524	0.552	0.561	0.553
1_2+ Styl.	0.601	0.602	0.601	0.576	0.594	0.566	0.652	0.661	0.651
1_3+ Styl.	0.583	0.583	0.583	0.543	0.564	0.529	0.597	0.603	0.596
1_4+ Styl.	0.566	0.566	0.566	0.557	0.553	0.542	0.608	0.608	0.604
2_3+ Styl.	0.575	0.576	0.575	0.549	0.576	0.547	0.615	0.626	0.616
2_4+ Styl.	0.57	0.571	0.57	0.577	0.581	0.572	0.61	0.617	0.611
3_4+ Styl.	0.538	0.538	0.537	0.546	0.544	0.53	0.552	0.569	0.554
1_2_3+ Styl.	0.599	0.599	0.599	0.565	0.587	0.561	0.64	0.648	0.639
1_2_4+ Styl.	0.578	0.578	0.578	0.578	0.586	0.576	0.645	0.65	0.644
1_3_4+ Styl.	0.595	0.595	0.595	0.556	0.558	0.547	0.6	0.603	0.597
2_3_4+ Styl.	0.554	0.555	0.554	0.58	0.588	0.579	0.594	0.605	0.596
1_2_3_4+ Styl.	0.592	0.593	0.592	0.589	0.589	0.589	0.624	0.631	0.624

**Table 8.** Classification results obtained by using five classes (highly positive-positive-neutral-negative-highly negative).

Table 8 shows that the SMO algorithm obtained better results than J48 and Naïve Bayes when five classes were involved. Individually, the second dimension (Psychological processes) provides the best result through the J48 and SMO algorithms. On the other hand, the Naïve Bayes algorithm obtained the best result through the first dimension (Linguistic processes). Conversely, the third and fourth dimensions provided the worst results. On the other hand, the combination of the first and second LIWC dimensions and the stylometric dimension provided the best classification results with an F-measure of 65.1%. In these results, the third and fourth dimensions are least revealing.

## 8 Discussion of Results

The classification results obtained show that reducing the number of classes the classifiers precision increases i.e. the classification with two categories (positive-negative) (see Fig. 3) provided better results than the classification with three (positive-neutral-negative) (see Fig. 4) and five (highly positive-positive-neutral-negative-highly negative) (see Fig. 5) categories. Thus, it is by virtue of the combination of fewer categories that the classification algorithm performs a better classification, probably due to the fact that in a bipolar system there is less space for the classification of slippery cases. It also means that additional criteria and features are required to get a fine-grained classification into 5 categories for instance.

With regard to the classification algorithms, SMO Algorithm for SVM classifier obtained the best results. This algorithm has proved to be simpler, easier to implement, generally faster (W. J. Han, 2015). Also, these results can be justified by the analysis presented in [40], where it is clearly shown how SVM models are more accurate in comparison to other classification algorithms such as: decision trees, neural network, Bayesian network, nearest neighbor. SVM algorithm has been successfully applied for text classification due to its main advantages: firstly, it has the ability to generalize well in high dimensional feature spaces; secondly, SVM eliminates the need for feature selection, making the application of text categorization considerably easier. Finally, another advantage of SVM algorithm over the conventional methods is its robustness [41].

Also, results showed that the combination of different LIWC dimensions provided better results than individual dimensions. Individually, the first and second LIWC dimensions (Linguistic processes and psychological processes) provided the best results. The first dimension contains grammatical words and the second dimension words related to the psychological process, such as: positive feelings, optimism, anxiety, sadness or depression, among others. Written opinions frequently contain a great amount of grammatical words and words related to emotional state of the author, which confirming the discriminatory potential of these dimensions in classification experiments. On the other hand, the

fourth dimension (personal concerns) provided the worst results, owing to the fact that the topic selected for this study, “Tourism”, bears little relation to the vocabulary corresponding to these categories. It can be stated that this dimension is the most content-dependent, and thus the least revealing. Also, general results show that the addition of the stylometric dimension improved the classification results. We ascribe this to the fact that most of the negative reviews contain more words than positive reviews, i.e. the number of words is an important feature to detect the polarity of the review.

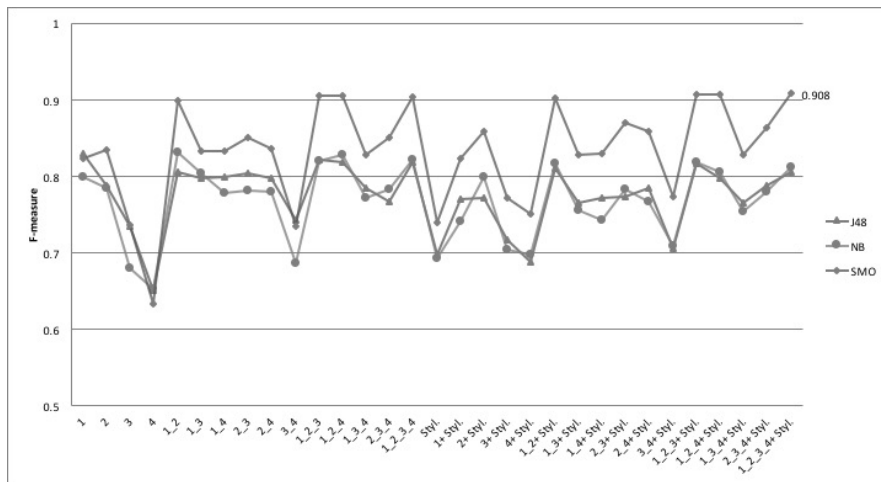
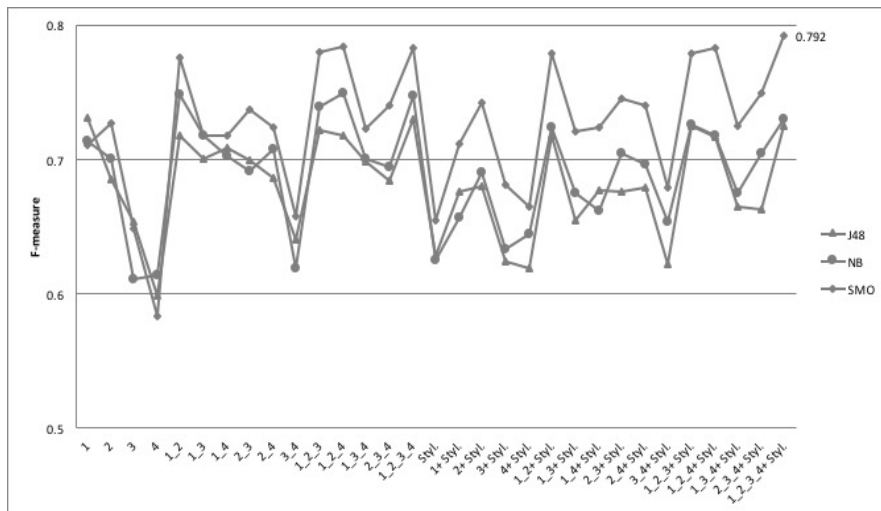
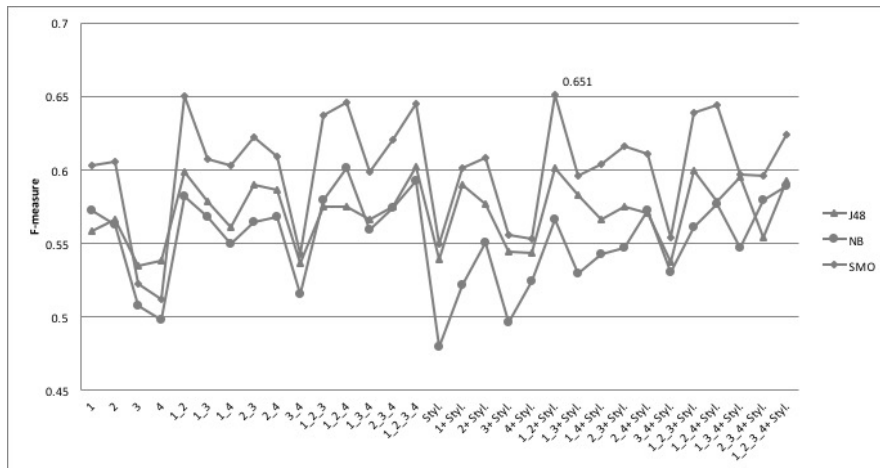


Fig. 3. Evaluation results for SMO, NB, J48, by using two classes.





**Fig. 4.** Evaluation results for SMO, NB, J48, by using three classes.



**Fig. 5.** Evaluation results for SMO, NB, J48, by using five classes.

## 8.1 Comparison

In order to compare our approach with other NLP tools used on sentiment analysis, the corpus used along our experiments was analyzed through the OpenNLP [42] and Stanford Classifier [43] tools. Then, the results obtained by both tools were compared with the results obtained by our approach.

The Stanford Classifier is a Java-based implementation of a maximum entropy classifier, which takes data and applies probabilistic classification. On the other hand, the Apache OpenNLP library is a machine learning based toolkit for the processing of natural language text. This includes a general Java maximum entropy package released under the GNU Lesser General Public License.

Table 9 shows that our proposal obtained encouraging results. Also, it obtained better results than Stanford classifier and OpenNLP, with an F-measure score of 90.8% for two classes, 79.2% for three classes, and 65.1% for five classes. We ascribe this to the following reasons: 1) the OpenNLP and Stanford NLP tools are based on maximum entropy models. Despite that Maximum entropy classifier provides good results in several proposals, SVM classifier outperforms the MaxEnt classifier in terms of the accuracy [44] [45] [46]. It can be justified for the SVM models advantages mentioned above in section “Discussion of results”, and 2) unlike of these tools, our approach is based on the psycholinguistic features

extraction to train the classifier. Therefore, we consider that these features are determinant in the classification results.

	Pos-Neg			Pos-Neu-Neg			highlyPos-Pos-Neu-Neg-highlyNeg		
	P	R	F1	P	R	F1	P	R	F1
Our approach	0.908	0.908	0.908	0.803	0.844	0.792	0.652	0.661	0.651
Stanford classifier	0.847	0.847	0.847	0.77	0.77	0.77	0.611	0.602	0.602
OpenNLP	0.853	0.849	0.848	0.782	0.75	0.728	0.647	0.65	0.646

**Table 9.** Comparison of Standford and OpenNLP with our proposal.

## 9 Conclusion and Future Work

In this piece of research, we presented the feasibility study of LIWC dimensions and stylometric dimension in the sentiment classification.

In order to conduct a comprehensive study, we considered two, three and five categories (positive-negative, positive-neutral-negative and highly positive-positive-neutral-negative-highly negative) for the classification of touristic reviews in Spanish language. Also, three classifiers (SVM, NB, and J48) were used in order to evaluate the efficacy of psychological and linguistic features for sentiment classification. The results showed that the classification of reviews with two categories “positive-negative” provides better results than with other categories. Also, we found that the best results were obtained with the SMO classifier.

Additionally, the findings reveal that the stylometric dimension improved the classification results. This fact indicates that these features are quite feasible for sentiment classification. Finally, we compared our approach with two natural NLP tools (OpenNLP and standford classifier) based on machine-learning. The performance was measured by using the precision, recall, and F-measure metrics. Our proposal obtained encouraging results with a high F-measure score of 90.8%.

With regard to future research, we consider to evaluate our approach by using new corpora concerning to different domains such as financial and movies domain. Also, we plan to verify the efficiency of our approach, i.e., the use of LIWC and stylometric dimensions on sentiment analysis, on non-Spanish languages such as French, English, Italian and German. Finally, we also attempt to apply this approach for detecting the satire in Twitter messages.

**Acknowledgments** This work has been partially supported by the Spanish Ministry of Economy and Competitiveness and the European Commission (FEDER / ERDF) through project KBS4FIA (TIN2016-76323-R). María del Pilar Salas-Zárate and Mario Andrés Paredes-Valverde are supported by the National Council of Science and Technology (CONACYT), and the Mexican government.

## References

- [1] M. Abdul-Mageed, M. Diab, and S. Kübler, ‘SAMAR: Subjectivity and sentiment analysis for Arabic social media’, *Comput. Speech Lang.*, vol. 28, no. 1, pp. 20–37, Jan. 2014.
- [2] S. Huang, Z. Niu, and C. Shi, ‘Automatic construction of domain-specific sentiment lexicon based on constrained label propagation’, *Knowl.-Based Syst.*, vol. 56, pp. 191–200, Jan. 2014.
- [3] A. Hogenboom, B. Heerschop, F. Frasincar, U. Kaymak, and F. de Jong, ‘Multi-lingual support for lexicon-based sentiment analysis guided by semantics’, *Decis. Support Syst.*, vol. 62, pp. 43–53, Jun. 2014.
- [4] Y. Bae and H. Lee, ‘Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers’, *J. Am. Soc. Inf. Sci. Technol.*, vol. 63, no. 12, pp. 2521–2535, Dec. 2012.
- [5] A. Montejó-Ráez, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. Ureña-López, ‘A knowledge-based approach for polarity classification in Twitter’, *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 2, pp. 414–425, Feb. 2014.
- [6] K. Singhal, B. Agrawal, and N. Mittal, ‘Modeling Indian General Elections: Sentiment Analysis of Political Twitter Data’, in *Information Systems Design and Intelligent Applications*, J. K. Mandal, S. C. Satapathy, M. K. Sanyal, P. P. Sarkar, and A. Mukhopadhyay, Eds. Springer India, 2015, pp. 469–477.
- [7] A. Duric and F. Song, ‘Feature selection for sentiment analysis based on content and syntax models’, *Decis. Support Syst.*, vol. 53, no. 4, pp. 704–711, Nov. 2012.
- [8] N. P. Cruz, M. Taboada, and R. Mitkov, ‘A machine-learning approach to negation and speculation detection for sentiment analysis’, *J. Assoc. Inf. Sci. Technol.*, p. n/a–n/a, Jun. 2015.
- [9] R. Moraes, J. F. Valiati, and W. P. GaviãoNeto, ‘Document-level sentiment classification: An empirical comparison between SVM and ANN’, *Expert Syst. Appl.*, vol. 40, no. 2, pp. 621–633, Feb. 2013.
- [10] R. Xia, F. Xu, J. Yu, Y. Qi, and E. Cambria, ‘Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis’, *Inf. Process. Manag.*, 2015.
- [11] Y. Liu, X. Yu, B. Liu, and Z. Chen, ‘Sentence-Level Sentiment Analysis in the Presence of Modalities’, in *Computational Linguistics and Intelligent Text Processing*, A. Gelbukh, Ed. Springer Berlin Heidelberg, 2014, pp. 1–16.
- [12] I. Peñalver-Martinez, F. Garcia-Sanchez, R. Valencia-Garcia, M. Á. Rodríguez-García, V. Moreno, A. Fraga, and J. L. Sánchez-Cervantes, ‘Feature-based opinion mining through ontologies’, *Expert Syst. Appl.*, vol. 41, no. 13, pp. 5995–6008, Oct. 2014.
- [13] A. Esuli and F. Sebastiani, ‘SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining’, in *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC’06, 2006*, pp. 417–422.
- [14] R. Valitutti, ‘WordNet-Affect: an Affective Extension of WordNet’, in *In Proceedings of the 4th International Conference on Language Resources and Evaluation, 2004*, pp. 1083–1086.

- [15] F. L. Cruz, J. A. Troyano, B. Pontes, and F. J. Ortega, 'ML-SentiCon: Un lexicón multilingüe de polaridades semánticas a nivel de lemas', *Procesamiento del Lenguaje Natural*, vol. 53, pp. 113–120, 2014.
- [16] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, 'Lexicon-Based Methods for Sentiment Analysis', *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, Apr. 2011.
- [17] M. Ghosh and K. Animesh, 'Unsupervised Linguistic Approach for Sentiment Classification from Online Reviews Using SentiWordNet 3.0.', *International Journal of Engineering Research and Technology*, vol. 2, no. 9, 2013.
- [18] V. Perez-Rosas, C. Banea, and M. Rada, 'Learning Sentiment Lexicons in Spanish', *LREC*, 2012.
- [19] S. Clematide and K. Manfred, 'Evaluation and extension of a polarity lexicon for German', presented at the Proceedings of the First Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, 2010, pp. 7–13.
- [20] I. Maks and P. Vossen, 'Different approaches to automatic polarity annotation at synset level', presented at the Proceedings of the First International Workshop on Lexical Resources, 2011, pp. 62–69.
- [21] A. Mageed, Muhammad, and M. Diab, 'Toward building a large-scale Arabic sentiment lexicon', presented at the Proceedings of the 6th International Global WordNet Conference, 2012, pp. 18–22.
- [22] I. Dehdarbehbahani, A. Shakery, and H. Faili, 'Semi-supervised word polarity identification in resource-lean languages', *Neural Netw.*, vol. 58, pp. 50–59, Oct. 2014.
- [23] M.-T. Martín-Valdivia, E. Martínez-Cámara, J.-M. Perea-Ortega, and L. A. Ureña-López, 'Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches', *Expert Syst. Appl.*, vol. 40, no. 10, pp. 3934–3942, Aug. 2013.
- [24] A. Balahur and M. Turchi, 'Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis', *Comput. Speech Lang.*, vol. 28, no. 1, pp. 56–75, Jan. 2014.
- [25] R. Hsu, B. See, and A. Wu, 'Machine Learning for Sentiment Analysis on the Experience Project', 2010.
- [26] P. P. B. Filho, T. A. Pardo, and S. M. Alusio, 'An evaluation of the brazilianportugueseliwc dictionary for sentiment analysis', presented at the In 9th Brazilian Symposium in Information and Human Language Technology, Fortaleza, Ceara, 2013.
- [27] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha, 'Comparing and Combining Sentiment Analysis Methods', in *Proceedings of the First ACM Conference on Online Social Networks*, New York, NY, USA, 2013, pp. 27–38.
- [28] C. J. Hutto and E. Gilbert, 'Vader: A parsimonious rule-based model for sentiment analysis of social media text', presented at the Eighth International AAAI Conference on Weblogs and Social Media, 2014.
- [29] M. del P. Salas-Zárate, E. López-López, R. Valencia-García, N. Aussenac-Gilles, Á. Almela, and G. Alor-Hernández, 'A study on LIWC categories for opinion mining in Spanish reviews', *J. Inf. Sci.*, vol. 40, no. 6, pp. 749–760, Dec. 2014.
- [30] Q. Ye, Z. Zhang, and R. Law, 'Sentiment classification of online reviews to travel destinations by supervised machine learning approaches', *Expert Syst. Appl.*, vol. 36, no. 3, Part 2, pp. 6527–6535, Apr. 2009.
- [31] G. Sidorov, S. Miranda-Jiménez, F. Viveros-Jiménez, A. Gelbukh, N. Castro-Sánchez, F. Velásquez, I. Díaz-Rangel, S. Suárez-Guerra, A. Treviño, and J. Gordon, 'Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets', in *Advances in Artificial Intelligence*, I. Batyrshin and M. G. Mendoza, Eds. Springer Berlin Heidelberg, 2013, pp. 1–14.
- [32] J. W. Pennebaker, T. J. Mayne, and M. E. Francis, 'Linguistic predictors of adaptive bereavement', *J. Pers. Soc. Psychol.*, vol. 72, no. 4, pp. 863–871, Apr. 1997.
- [33] M. E. Francis and J. W. Pennebaker, 'LIWC: Linguistic Inquiry and Word Count', Dallas, TX: Southern Methodist University, 1993.

- [34] J. W. Pennebaker, M. E. Francis, and R. J. Booth, 'Linguistic Inquiry and Word Count', Mahway: Lawrence Erlbaum Associates, vol. 71, 2001.
- [35] N. Ramírez-Esparza, J. W. Pennebaker, F. A. García, and R. Suriá Martínez, 'La psicología del uso de las palabras: un programa de computadora que analiza textos en español', *Thepsychology of word use: a computerprogramthatanalyzestexts in Spanish*, Jun. 2007.
- [36] M. Rushdi Saleh, M. T. Martín-Valdivia, A. Montejo-Ráez, and L. A. Ureña-López, 'Experiments with SVM to classify opinions in different domains', *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14799–14804, Nov. 2011.
- [37] A. Montejo-Ráez, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. Ureña-López, 'Ranked WordNet graph for Sentiment Polarity Classification in Twitter', *Comput. Speech Lang.*, vol. 28, no. 1, pp. 93–107, Jan. 2014.
- [38] T. Chalothom and J. Ellman, 'Simple Approaches of Sentiment Analysis via Ensemble Learning', in *Information Science and Applications*, K. J. Kim, Ed. Springer Berlin Heidelberg, 2015, pp. 631–639.
- [39] R. R. Bouckaert, E. Frank, M. A. Hall, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, 'WEKA—Experiences with a Java Open-Source Project', *J. Mach. Learn. Res.*, vol. 11, p. 2533–2541, Sep. 2010.
- [40] H. Bhavsar and A. Ganatra, 'A Comparative Study of Training Algorithms for Supervised Machine Learning', *International Journal of Soft Computing and Engineering*, vol. 2, no. 4, pp. 74–81, 2012.
- [41] N. Deng, Y. Tian, and C. Zhang, *Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions*. CRC Press, 2012.
- [42] J. Baldrige, 'the opennlp project', openNLP, 2010. [Online]. Available: <https://opennlp.apache.org/>. [Accessed: 18-May-2015].
- [43] B. MacCartney, 'Stanford Classifier', The Stanford Natural Language Processing Group. [Online]. Available: <http://nlp.stanford.edu/software/classifier.shtml>. [Accessed: 18-May-2015].
- [44] M. Anjaria and R. M. R. Guddeti, 'Influence factor based opinion mining of Twitter data using supervised learning', in *2014 Sixth International Conference on Communication Systems and Networks (COMSNETS)*, 2014, pp. 1–8.
- [45] N. T. Duyen, N. X. Bach, and T. M. Phuong, 'An empirical study on sentiment analysis for Vietnamese', in *2014 International Conference on Advanced Technologies for Communications (ATC)*, 2014, pp. 309–314.
- [46] S. Chinthala, R. Mande, S. Manne, and S. Vemuri, 'Sentiment Analysis on Twitter Streaming Data', in *Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1*, S. C. Satapathy, A. Govardhan, K. S. Raju, and J. K. Mandal, Eds. Springer International Publishing, 2015, pp. 161–168.